

Personal Identification Based on Mobile-based Keystroke Dynamics

Alaa Tharwat^{1,2,6}, Abdelhameed Ibrahim^{3,6}, Tarek Gaber^{4,6}, Aboul Ella Hassanien^{5,6}

¹ Faculty of Engineering, Suez Canal University, Egypt

² Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, 60318 Frankfurt am Main, Germany

³ Faculty of Engineering, Mansoura University, Egypt

⁴ Faculty of Computers and Informatics, Suez Canal University, Egypt

⁵ Faculty of Computers and Information, Cairo University, Egypt

⁶ Scientific Research Group in Egypt (SRGE) <http://www.egyptscience.net>

Abstract. This paper is addressing the personal identification problem by using mobile-based keystroke dynamics of touch mobile phone. The proposed approach consists of two main phases, namely feature selection and classification. The most important features are selected using Genetic Algorithm (GA). Moreover, Bagging classifier used the selected features to identify persons by matching the features of the unknown person with the labeled features. The outputs of all Bagging classifiers are fused to determine the final decision. In this experiment, a keystroke dynamics database for touch mobile phones is used. The database, which consists of four sets of features, is collected from 51 individuals and consists of 985 samples collected from males and females with different ages. The results of the proposed model conclude that the third subset of features achieved the best accuracy while the second subset achieved the worst accuracy. Moreover, the fusion of all classifiers of all ensembles will improve the accuracy and achieved results better than the individual classifiers and individual ensembles.

Keywords: Keystroke dynamics, RHU keystroke, Genetic algorithm, Bagging classifier, Personal identification

1 Introduction

The number of smartphone users in 2016 was 2.1 billion, and it is expected to reach around 2.5 billion in 2019 [1]. Along with sensing capabilities, the pervasive feature of smartphones has changed the prospective of people's everyday life. Smartphones are now full of personal information including bank account details, medical information (e.g., sugar level and heart rate), and other personal credentials required for different services and applications. In such context, smartphone's users become worried about the confidentiality and privacy of their data and information [2]. Thus, maintaining the privacy of such users' sensitive data and the information is a very vital issue.

Passwords, personal identification number (PIN), and pattern lock are considered the most common methods to protect the identity of smartphone's owners from illegal accesses. However, the PIN method is subject to the dictionary attack as well as shoulder surfing attack. Also, Pattern locks are subject to side channel attack. The most importantly, such commonly used methods for identifying a legitimate owner of a smartphone fail to recognize and detect an adversary when she/he has broken the PIN/Password [3] making these methods ineffective for continuous authentication.

Continuous authentication is a process in which user's behavioral biometrics are used for authenticating a smartphone user [4]. Mobile authentication based on behavioral biometrics depends on learning from user behaviors which do not change over a period. This includes gait patterns [5], signature [6], voice [7], and touchscreen interactions [8, 9]. Such characteristics are then implicitly used to identify a given smartphone user thus preventing an illegitimate user from accessing the device. The authentication decision in this context is taken based on distinctive features collected from the user's behaviors.

This paper describes an approach addressing the personal identification problem by using mobile-based keystroke dynamics. In the proposed approach, a keystroke dynamics dataset is used [10]. The most important features are selected using GA. Moreover, Bagging classifier used the selected features to identify persons by matching the features of the unknown person with the labeled features.

The rest of the paper is organized as follows; Section (2) presents the feature selection technique using a genetic algorithm and the Bagging classifier ensemble. Our proposed approach are presented in Section (3) while, the experimental results are given in Section (4). The discussion of the obtained results and the conclusion of the paper are presented in Section (5) and (6), respectively.

2 Preliminaries

2.1 Genetic Algorithm (GA)

Genetic Algorithm (GA) is one of the well-known optimization algorithms. The main idea of GA depends on evaluating and modifying the current population, i.e., a set of solutions, which are represented in binary numbers and selects the best solutions to generate new solutions from it. Hence, GA searches mainly in binary spaces to find the optimal solutions [11].

Mathematically, given the initial population set $P = \{p_1, p_2, \dots, p_n\}$, where p_i is one candidate, agent or chromosome. Each chromosome consists of many genes and the value of each gene is zero or one. The population in GA is modified to create a new generation in each iteration. This modification includes three main processes, namely, (1) crossover, (2) mutation, and (3) selection. In the crossover process, the information of two chromosomes, i.e., parents, are exchanged to generate two new chromosomes, i.e., children. There are three methods of crossover between chromosomes: (1) single point, (2) multi-points,

and (3) uniform. For example, in the single point crossover method, the information of two chromosomes 00100101 and 10111010 are exchanged and this can be achieved by selecting one cutting point and exchange one part from each parent to generate new chromosomes, 00101010 and 10110101. The cutting point is chosen randomly in many algorithms. In the mutation process, the original chromosome, i.e., parent, mutate randomly by changing the value of one or more bits to generate a new chromosome, i.e., child. Moreover, GA allows the chromosomes to mutate and generate a near identical chromosome from the original one. The current chromosomes and the new generated ones are evaluated using the fitness function and the best chromosomes are selected (survived) to the next generation, this is called selection step [12].

The performance of GA is controlled by many parameters such as the population size (N), crossover rate (P_c), mutation rate (P_m), and the number of iterations (T). The population size represents the total number of chromosomes while the crossover rate represents the probability of accepting an eligible pair of chromosomes for crossover. The mutation rate is the probability of switching bits in the chromosomes [13]. The steps of GA are summarized in Algorithm 1.

Algorithm 1 Genetic Algorithm

```

1: Initialize population  $P = \{p_1, p_2, \dots, p_n\}$ , where  $n$  represents the number of chromosomes in the populations, and the total number of generations is  $T$ .
2: for all (Generations  $t < T$ ) do
3:   for all ( $i = 1$  to  $n/2$ ) do
4:     Select two parents  $p_i$  and  $p_{n/2+i}$  from  $P$ .
5:     Offspring  $O_i = \text{CrossOver}(p_i, p_{n/2+i})$ 
6:     Offspring  $O_i = \text{Mutation}(\text{Offspring } O_i)$ 
7:   end for
8:   Combine the original population  $P$  and the new offsprings  $O_i$  into new  $P_{New} = P + O_i$ 
9:   Evaluate all chromosomes in the new population  $P_{New}$ 
10:  Select the best chromosomes to survive for the next generation
11: end for

```

2.2 Bagging Classifiers

Bagging is an acronym from *Bootstrap AGGregatING* and it is one of the well-known ensemble methods that creates its ensemble by training different classifiers on a random distribution of the training set. Many samples from the original training set (X) are drawn randomly with replacement to form a new training set X_{New} , which is then used for training one individual classifier C_i in the Bagging ensemble. The accuracy of the classifiers of each pattern is chosen with equal probability; i.e., not weighted as in AdaBoost [14], then combine the outputs of all classifiers using uniform averaging or voting over class labels. Since the Bagging resamples the training set with replacement, so some samples are

chosen multiple times, while the other samples are left out [15, 16]. The details of Bagging classifier are in Algorithm [2].

Algorithm 2 Bagging Algorithm

Input: Required ensemble size T , Training set $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

while ($t < T$) **do**

Build a dataset S_t , by randomly sampling N times from S (with replacement)

Train a model h_t using S_t and add it to the ensemble

end while

for all (new test pattern (\hat{x}, \hat{y})) **do**

for all (models h_t) **do**

Classify the unknown or test pattern using the model h_t , the output of the classification is C_t

end for

Combine the outputs of the model $C_i, i = 1, 2, \dots, T$

end for

3 Proposed Model

The proposed approach in this research consists of two main phases, namely feature selection and classification. As shown in Fig. (1), our dataset that is used in this research has four different sets (f_1, f_2, f_3 , and f_4) of features. In the first phase, GA is applied for each feature subset separately to select the best features in each subset ($f_{new1}, f_{new2}, f_{new3}$, and f_{new4}). In the second phase, the data were partitioned into training set which is used to train the Bagging model and testing set which is used to evaluate the proposed model. More details about each phase in the next two sections.

3.1 Feature Selection Phase

In this phase, the GA was employed for selecting a small subset d from a large set of features D to remove independent and redundant features. The large set of features is denoted by, $X = \{f_1, f_2, \dots, f_D\}$. While the small subset of features denoted by, $Y = \{f_1, \dots, f_d\}$, where $d \leq D$. GA first initializes a set of chromosomes coded in binary code randomly. Each chromosome represents a subset of features, and each bit of the chromosome denoting the presence, i.e., 1, or absence, i.e., 0, of that feature in this subset. Thus, the dimension of search space R^d for the features is determined by the total number of features d [17].

In GA, the fitness function should ensure that the fitter chromosome has a high probability to survive for the next generations. But, in feature selection application, the goal of the fitness function of the GA is to evaluate a subset of features requires an estimation of the misclassification rate for that subset.

3.2 Classification Phase

In this phase, the new subset of features ($f_{newi}, i = 1, 2, 3, 4$) that were selected in the first phase are used. This phase consists of two main steps; training and testing. Each new subset of features is divided into training and testing sets. In the training step as shown in Fig. (1), each training set from the new subsets of features is used to train one Bagging model. Thus, four different models are built using the new four feature subsets that were selected. In the testing phase, each testing subset is matched or classified with their corresponding training model that is built using the corresponding feature subset. The outputs of the four classification processes are combined using majority voting to get the final decision.

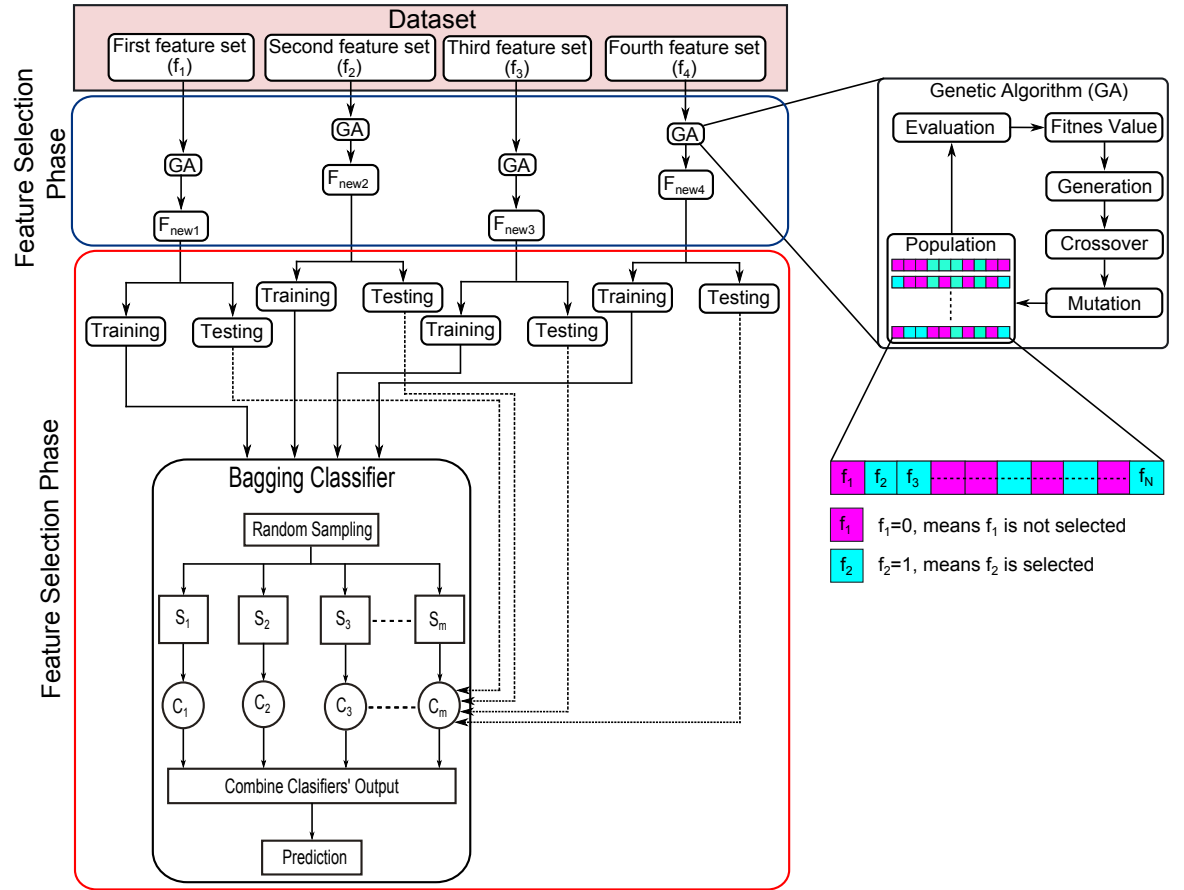


Fig. 1: Block diagram of the proposed model.

4 Experimental Results

4.1 Dataset

In this experiment, a keystroke dynamics database for touch mobile phones was used. The database is collected from 51 individuals. Each individual is requested to type a password "rhu.university" 15 times during three different sessions [10]. The data in each session are collected separately and the average time between each session was five days. The database consists of 985 samples collected from males and females with different ages as shown in Fig. (2). The distribution of the database is shown in Table (1) and Fig. (2).

Table 1: Distribution of the keystroke dynamics database.

Gender	Age Group						Total	
	7-18		19-29		30-65			
	Users	Samples	Users	Samples	Users	Samples	Users	Samples
Male	7	137	16	281	3	59	26	477
Female	4	75	14	263	7	140	25	478
Total	11	212	30	544	10	199	51	955

In the database there are four information were extracted from each user includes the timing between a key pressure and a key release, timing between a key release and a key pressure, the timing between two key pressures, and timing between two key releases [10].

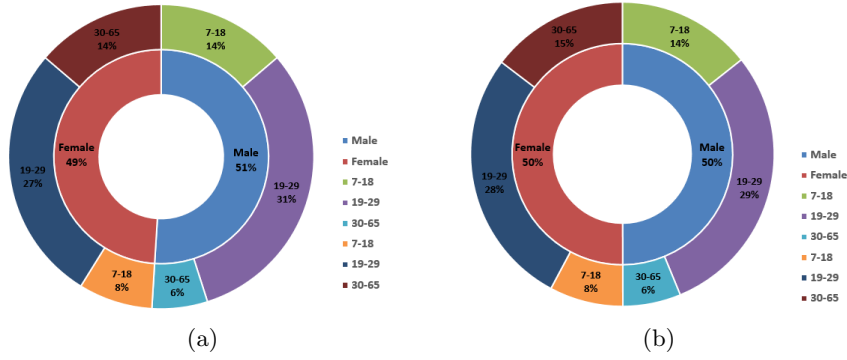


Fig. 2: A distribution of samples and users of dynamic keystrokes database, (a) distribution of the users over gender and age groups, (b) distribution of the samples over gender and age group.

In this research, two different experiments are used to test our proposed approach. The goal of the first experiment was to select the most important

features using GA. While in the second experiment, the Bagging classifier was used to match the testing and training patterns.

4.2 Feature Selection Experiment

This experiment is conducted to select the most powerful features. As mentioned above, the dataset that was used in this research is consists of four subsets of features. In this experiment, GA is used to select the best subset of features from each subset. Each subset from the original features was used as an input to the GA algorithm, and the GA selects the best subset of features. The GA has many parameters that affect the performance of GA. In this experiment, the population size of GA is 50, crossover rate=0.6, mutation rate=0.01, and the number of iterations=45. The results of this scenario are summarized in Table (2) and Fig. (3).

Table 2: The selected features (F_{newi}) using GA algorithm

Feature set	Selected Features
F_{new1}	{2,3,4,5,6,7,8,10,11,13}
F_{new2}	{2,4,6,8,10,11,13,14}
F_{new3}	{1,3,4,5,6,7,8,9,10,13}
F_{new4}	{1,3,6,7,8,9,10,11,12,13}

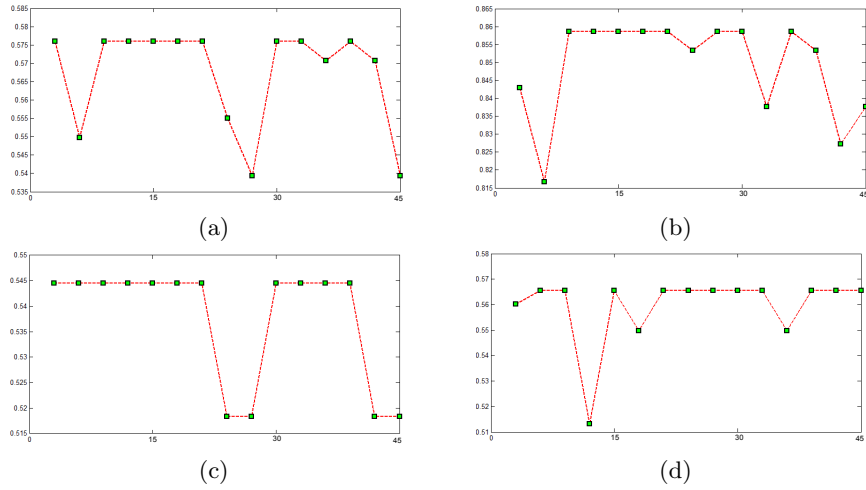


Fig. 3: Fitness values vs. generation number, (a) First feature subset, (b) second feature subset, (c) third feature subset, (d) fourth feature subset. chips.

4.3 Classification Experiment

In this experiment, the second phase of the proposed model shown in Fig. (1) is investigated. This experiment consists of two steps. In the first step, four different Bagging classifiers are trained. Each Bagging classifier has five weak or individual decision tree classifiers. Each new features subset was divided into training and testing subsets as follows, ($f_{newi} = TR_{newi} + TS_{newi}$). Each subset of the training features ($TR_{newi}, i = 1, 2, 3, 4$) represents the input to one Bagging classifier. Each Bagging classifier generates a model, which will be used in the testing step to classify the unknown or test samples. The results of the training step are summarized in Table (3). In the second step, the testing samples were classified using the Bagging models that were calculated in the training step. Each testing subset was matched separately and the outputs of all Bagging classifiers were fused into one decision. The results of the testing step are summarized in Fig. (4).

Table 3: Accuracy (in %) of the individual classifiers of the Bagging ensemble, the accuracy of the ensemble, and the fusion of the four ensembles in the training step.

Classifier	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Classifier 1	97.3	78.3	92.9	91.5
Classifier 2	92.4	76.3	92.1	90.8
Classifier 3	90.7	77	92.8	91.4
Classifier 4	92.1	78.4	91.4	94.7
Classifier 5	92.6	74.9	92.9	93.1
Bagging Accuracy	100	99.2	100	100
Fusion of All Ensembles	100			

Table 4: Accuracy (in %) of the individual classifiers of the Bagging ensemble, the accuracy of the ensemble, and the fusion of the four ensembles in the testing step.

Classifier	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4
Classifier 1	58.7	37.3	55.9	56.4
Classifier 2	57.7	41.5	59.5	57.2
Classifier 3	53.5	39.9	60.6	59.2
Classifier 4	53	40.2	59.5	56.9
Classifier 5	52.2	40.4	57.2	55.6
Bagging Accuracy	63.4	41	70.7	68.6
Fusion of All Ensembles	83.8			

5 Discussions

From Table (2) many remarks can be seen. The number of selected features in all subsets was ten, except the second feature subset consists of eight features, which reflects that the second feature subset has redundant features more than the other feature subsets. From Fig. (3) we note that the fitness function fluctuated up and down and in the first and third subsets of features the fitness function value reached its minimum. In our experiment, in all runs (i.e. all feature subsets) the best solution is used to select the new subset of features.

Table (3) shows the experimental results of the training step in the classification phase and it has many findings. Firstly, the accuracy results of the individual classifiers in all subset of features are more than 90%, except in the second subset achieved accuracy ranged from 74.9% to 78.4%, which reflects that the second subset of features is not robust enough again identifying persons. Secondly, the first subset of features achieved the best accuracies. Thirdly, the accuracy of each Bagging, which represents the fusion of all individual classifiers in the Bagging classifier achieved an excellent accuracy ranged from 99.2% to 100%, which reflects that the individual classifiers are independent; and hence, the fusion rule (majority voting) achieved good results. The fifth remark is that the fusion of all Bagging achieved accuracy equal to 100%.

Table (4) presents the results of the testing step in the classification phase. As shown in the table, the third subset of the selected features achieved the best results while the second subset achieved the minimum accuracy. The accuracy of the Bagging classifier achieved accuracy better than all its individual classifiers, and it was ranged from 41% to 70.7%, while the fusion of all Bagging classifiers was 83.8%.

6 Conclusions and Future Work

In this study, a proposed approach is used to identify persons based on keystroke dynamics of touch mobile phone. A dataset of keystroke dynamics, which consists of four sets of features, is used. The proposed approach consists of two phases, namely, feature selection and classification. In the first phase, GA is used to select the best subset of features from each set of features. While, in the second phase, the features that were selected in the first phase are used to identify persons using Bagging classifier. The unknown person is identified based on each feature subset separately using an ensemble of Bagging classifiers. The outputs of all Bagging classifiers are fused to determine the final decision. The results of the proposed model conclude that the specific combination of the features, i.e., timing between a key pressure and a key release, timing between a key release and a key pressure, the timing between two key pressures, and timing between two key releases achieved the best accuracy while the second subset achieved the worst accuracy. Moreover, the fusion of all classifiers of all ensembles will improve the accuracy and achieved results better than the individual classifiers and individual ensembles. Future work includes, using different feature

selection methods and also investigate other ensemble classifiers to increase the performance of the approach.

References

1. www.statista.com: Smartphone users worldwide 2014-2020 ((accessed October 11, 2017))
2. Baljit Singh Saini, N.K., Bhatia, K.S.: Keystroke dynamics for mobile phones: A survey. *Indian Journal of Science and Technology* **9**(6) (2016) 1–8
3. Md Liakat Ali, John V. Monaco, C.C.T., Qiu, M.: Keystroke biometric systems for user authentication. *Journal of Signal Processing Systems* **86**(2–3) (2017) 175–190
4. Alzubaidi, A., Kalita, J.: Authentication of smartphone users using behavioral biometrics. *IEEE Communications Surveys & Tutorials* **18**(3) (2016) 1998–2026
5. Gafurov, D.: A survey of biometric gait recognition: Approaches, security and challenges. In: *Annual Norwegian computer science conference*. (2007) 19–21
6. Blanco-Gonzalo, R., Miguel-Hurtado, O., Mendaza-Ormaza, A., Sanchez-Reillo, R.: Handwritten signature recognition in mobile scenarios: Performance evaluation. In: *Security Technology (ICCST), 2012 IEEE International Carnahan Conference on*, IEEE (2012) 174–179
7. Lu, H., Bernheim Brush, A., Priyantha, B., Karlson, A., Liu, J.: Speakersense: Energy efficient unobtrusive speaker identification on mobile phones. *Pervasive Computing* (2011) 188–205
8. Zheng, N., Bai, K., Huang, H., Wang, H.: You are how you touch: User verification on smartphones via tapping behaviors. In: *Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, IEEE (2014) 221–232
9. Buza, K.: Person identification based on keystroke dynamics: Demo and open challenge. In: *CAiSE Forum, 28th International Conference on Advanced Information Systems Engineering*. (2016)
10. El-Abed, M., Dafer, M., Khayat, R.E.: Rhu keystroke: A mobile-based benchmark for keystroke dynamics systems. In: *International Carnahan Conference on Security Technology (ICCST), 2014, IEEE* (2014) 1–4
11. Diaz, J.M., Pinon, R.C., Solano, G.: Lung cancer classification using genetic algorithm to optimize prediction models. In: *The 5th International Conference on Information, Intelligence, Systems and Applications (IISA,2014)*, IEEE (2014) 1–6
12. Li, S., Wu, H., Wan, D., Zhu, J.: An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems* **24**(1) (2011) 40–48
13. Grefenstette, J.J.: *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*. Psychology Press (2013)
14. Gaber, T., Tharwat, A., Hassanien, A.E., Snasel, V.: Biometric cattle identification approach based on webers local descriptor and adaboost classifier. *Computers and Electronics in Agriculture* **122** (2016) 55–66
15. Breiman, L.: Bagging predictors. *Journal of Machine learning* **24**(2) (1996) 123–140
16. Tharwat, A., Gaber, T., Awad, Y.M., Dey, N., Hassanien, A.E.: Plants identification using feature fusion technique and bagging classifier. In: *The 1st International Conference on Advanced Intelligent System and Informatics (AISII2015)*, November 28-30, 2015, Beni Suef, Egypt, Springer (2016) 461–471
17. Wojciech, S., Jack, S.: A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters* **10**(5) (1989) 335–347